■ ปรีดา เลิศพงศ์วิภูษณะ และ
สมชาย นำประเสริฐชัย

คุยเฟื่องเรื่อง Linux : เว็บเซิร์ฟเวอร์

เว็บเซิร์ฟเวอร์ที่นิยมใช้กันมาก บนระบบปฏิบัติการหลายประเภท

ก่อนอื่นต้องขออภัยกับท่านผู้อ่านด้วยนะครับที่เรื่องราวของ Linux นั้นขาดหายไปจากวารสารไมโครคอมพิวเตอร์ในเดือนที่ผ่าน ก็ไม่มีอะไรเป็นพิเศษหรอกครับ มัวแต่เพลินกับวันหยุดในเดือนเมษายนนั่นเอง มารู้ตัวอีกครั้งก็ปิดรับต้นฉบับไปแล้ว ช่วงนี้ก็เป็นช่วงเปิดเทอม สถาบันการศึกษาต่าง ๆ ก็รู้สึกคึกคัก มีชีวิตชีวาขึ้น หลังจากเงียบเหงามานานในช่วงปิดเทอม สำหรับเรื่องราวของ Linux ในฉบับนี้เป็นเรื่องของ อาปาเช่ ซึ่งเป็นเว็บเซิร์ฟเวอร์ที่นิยมกันในขณะนี้จากการสำรวจของ Netcraft ในเดือนมกราคม ต้นปีที่ผ่านมานี้พบว่า มีการใช้อาปาเช่เป็นเว็บเซิร์ฟเวอร์มากกว่า 53 เปอร์เซ็นต์

เว็บเซิร์ฟเวอร์คืออะไร

เว็บเซิร์ฟเวอร์คือโปรแกรมที่ทำหน้าที่ให้บริการข้อมูลเอกสารกับระบบ World Wide Web (WWW) โดยจะคอยรับการร้องขอเอกสารจากเว็บเบราเซอร์ เช่น NetScape Communicator หรือ Internet Explorer สำหรับการติดต่อระหว่างเว็บเซิร์ฟเวอร์กับเว็บเบราเซอร์จะต้องใช้โปรโตคอล HyperText Transfer Protocol (HTTP) ในการขนย้ายข้อมูลเอกสารและสถานะของเอกสารต่าง ๆ โดยโปรโตคอลนี้จะสามารถใช้งานได้เฉพาะบนระบบเครือข่ายที่ใช้โปรโตคอลทีซีพี/ไอพี (TCP/IP) เท่านั้น เอกสารที่จะใช้ในการบริการของระบบเว็บเซิร์ฟเวอร์จะสามารถจัดเก็บได้ในหลาย ๆ รูปแบบ แต่รูปแบบที่นำมาซึ่งประสิทธิภาพสูงสุดในการนำเสนอจะต้องถูกจัดเก็บอยู่ในรูปแบบของภาษา HyperText Markup Language (HTML) ซึ่งเป็นรูปแบบที่สามารถนำเสนอข้อมูลข่าวสารได้หลากหลาย เช่น ภาพ เสียง ภาพยนตร์ และการโต้ตอบกับผู้ใช้ (Interactive) ในปัจจุบันนี้ HTML มีด้วยกันหลายรุ่น เช่น HyperText Markup Language (HTML), Standard Markup Language (SGML), Extended Markup Language (XML)

เมื่อเว็บเบราเซอร์ทำการร้องขอเอกสารที่จัดเก็บไว้ในรูปแบบ HTML ไปยังเว็บเซิร์ฟเวอร์ ส่วนของเว็บเซิร์ฟเวอร์จะทำการค้นหาและจัดส่งเอกสารกลับไปยังเว็บเบราเซอร์ เมื่อเว็บเบราเซอร์ได้รับเอกสารทั้งหมดแล้วจะทำการแปลภาษา HTML แล้วนำเสนอเอกสารในรูปแบบกราฟิก

Linux

เว็บเซิร์ฟเวอร์ต่าง ๆ ที่มีการใช้งาน

Apache เป็นเว็บเซิร์ฟเวอร์ของบริษัท Apache Group. ที่ทำงานบนระบบปฏิบัติการของ Microsoft Windows NT ระบบปฏิบัติการลีนุกซ์ และอื่น ๆ

Microsoft Internet Information Server เป็นเว็บเซิร์ฟเวอร์ของบริษัทไมโครซอฟท์ ที่ทำงานบนระบบปฏิบัติการของ Microsoft Windows NT ปัจจุบันมีถึงรุ่นที่ 4

WebSite เป็นเว็บเซิร์ฟเวอร์ของบริษัท O'Reilly & Associates ltd. ที่ทำงานบนระบบปฏิบัติการของ Microsoft Windows NT และระบบปฏิบัติการลีนุกซ์

ทำไมต้องอาปาเช่เว็บเซิร์ฟเวอร์

อาปาเช่เว็บเซิร์ฟเวอร์เป็นโปรแกรมแจกฟรี ไม่ต้องเสียค่าใช้จ่ายในการใช้ และมีการดูแลปรับปรุงแก้ไขข้อผิดพลาดต่าง ๆ ที่เกิดขึ้นอย่างต่อเนื่องและรวดเร็ว เช่นเดียวกันกับระบบปฏิบัติการลีนุกซ์ มีมาให้พร้อมกับชุดติดตั้งระบบปฏิบัติการลีนุกซ์ของหลาย ๆ บริษัท หรือถ้าต้องการโปรแกรมรุ่นใหม่ล่าสุดก็สามารถหาสำเนาได้จากเอฟทีพีเซิร์ฟเวอร์ต่าง ๆ มากมาย และนอกจากนี้ยังมีการเปิดเว็บไซท์ให้บริการข้อมูลข่าวสารต่าง ๆ ที่ http://www.apache.org มีข้อมูลทั้งการติดตั้ง การดูแลและจัดการ การแก้ไขปัญหาที่เกิดขึ้นกับอาปาเช่เว็บเซิร์ฟเวอร์ มีการทดสอบประสิทธิภาพในการให้บริการเมื่อเทียบกับเว็บเซิร์ฟเวอร์อื่น ๆ นอกจากนี้แล้วอาปาเช่เว็บเซิร์ฟเวอร์ยังมีรุ่นที่สามารถทำงานบนระบบปฏิบัติการอื่น ๆ ได้อีกด้วย

อาปาเช่เว็บเซิร์ฟเวอร์

ชื่อเรียกอาปาเช่ "Apache" นั้นมาจากคำว่า "A PAtCHy server" ซึ่งเป็นชื่อเฉพาะนะครับ หากถ้าเป็นคำว่า apache นั้นก็ความหมายเป็นอีกแบบนะครับ อย่างในพจนานุกรมอิเล็กทรอนิกส์ของ ส.เศรษฐบุตร ก็ให้ความหมายว่าเป็นกุ๊ยในปารีส การที่อาปาเช่มีชื่อเช่นนี้ก็เนื่องจากอาปาเช่เว็บเซิร์ฟเวอร์ในรุ่นแรก ๆ เป็นการติดตั้งเว็บเซิร์ฟเวอร์ NCSA แล้วทำการแก้ไขโปรแกรมต่าง ๆ ด้วยชุดไฟล์แก้ไข (Patch Files) ที่ประกอบด้วยไฟล์ต่าง ๆ หลังจากทำการดำเนินการแล้วถึงจะได้อาปาเช่เว็บเซิร์ฟเวอร์ อาปาเช่เว็บเซิร์ฟเวอร์ในปัจจุบันได้มาถึงรุ่นที่ 1.3

ในปี พ.ศ. 2538 อาปาเช่เว็บเซิร์ฟเวอร์ได้ถูกพัฒนาขึ้น โดยได้นำต้นแบบมาจาก โปรแกรมเว็บเซิร์ฟเวอร์ของ NCSA รุ่นที่ 1.3 เป็นจุดเริ่มต้นของการพัฒนาโปรแกรมตัวนี้ หลังจากนั้นอีก 2 ปี ในปี พ.ศ. 2540 อาปาเช่เว็บเซิร์ฟเวอร์ก็ถูกสร้างขึ้นใหม่หมดและพัฒนาจนสมบูรณ์ จนเป็นที่ยอมรับและนิยมใช้งานกันอย่างมากมายบนเครือข่ายอินเทอร์เน็ต ด้วยเหตุผลว่าเป็นเว็บเซิร์ฟเวอร์ที่มีประสิทธิภาพและความเร็วสูงในการให้บริการ และมีความสามารถใหม่ ๆ ให้ใช้มากมาย จนถึงปี พ.ศ. 2541 มีผู้ตัดสินใจในการติดตั้งและใช้งานไปแล้วจำนวน 1.2 ล้านเซิร์ฟเวอร์ทั่วโลก

การติดตั้งอาปาเช่เว็บเซิร์ฟเวอร์

วิธีการติดตั้งอาปาเช่เว็บเซิร์ฟเวอร์จะมีด้วยกันหลายวิธีคือ การติดตั้งอาปาเช่เว็บเซิร์ฟเวอร์พร้อมระบบปฏิบัติการลีนุกซ์ หรือติดตั้งอาปาเช่เว็บเซิร์ฟเวอร์หลังจากได้ทำการติดตั้งระบบปฏิบัติแล้วก็ได้

การดูแลจัดการอาปาเช่เว็บเซิร์ฟเวอร์

หลังจากได้ทำการติดตั้งอาปาเช่เว็บเซิร์ฟเวอร์เป็นที่เรียบร้อยแล้ว จะพบว่าอาปาเช่เว็บเซิร์ฟเวอร์จะมีไฟล์ที่สำคัญ ๆ ที่ใช้กำหนดคุณสมบัติต่าง ๆ โดยปกติไฟล์เหล่านี้จะถูกเก็บไว้ที่ไดเรคทอรี /etc/httpd/conf ซึ่งมีดังต่อไปนี้

httpd.conf เป็นไฟล์ที่ใช้เก็บข้อกำหนดเพื่อควบคุมคุณสมบัติการทำงานต่าง ๆ ของอาปาเช่เว็บเซิร์ฟเวอร์

access.conf เป็นไฟล์ที่ใช้เก็บข้อกำหนดเพื่อควบคุมการเข้าถึงของผู้ใช้และเอกสารต่าง ๆ ของอาปาเช่เว็บเซิร์ฟเวอร์ ถ้าไม่มีการกำหนดใด ๆ จะถือว่าเป็นการอนุญาตให้เข้าถึงเอกสารต่าง ๆ ได้ทั้งหมดจากผู้ใช้ทุก ๆ คน

srm.conf เป็นไฟล์ที่ใช้เก็บข้อกำหนดเพื่อระบุถึงทรัพยากรต่าง ๆ ที่จำเป็นของอาปาเช่เว็บเซิร์ฟเวอร์ เช่น ไดเรคทอรีของเอกสาร ไดเรคทอรีของโปรแกรมต่าง ๆ

การกำหนดคุณสมบัติไฟล์ httpd.conf

การกำหนดลักษณะการทำงานและการให้บริการของอาปาเช่เว็บเซิร์ฟเวอร์จะประกอบด้วยข้อกำหนดหรือที่เรียกว่าไดเรคทีฟ (Directive) ดังต่อไปนี้

- **ServerType {inetd, standalone}**

ServerType เป็นการกำหนดรูปแบบการให้บริการของเว็บเซิร์ฟเวอร์ มีตัวเลือกด้วยกันดังต่อไปนี้

inetd คือการให้เว็บเซิร์ฟเวอร์ทำงานภายใต้การควบคุมของโปรแกรม inetd อีกขั้นหนึ่ง โดยเมื่อเริ่มต้นโปรแกรม inetd จะทำหน้าที่ในการตรวจสอบพอร์ตที่ใช้ในการบริการของเว็บเซิร์ฟเวอร์ว่ามีการขอบริการหรือไม่ หากมีการติดต่อขอบริการโปรแกรม inetd จะสั่งให้เว็บเซิร์ฟเวอร์เริ่มต้นทำงานทันทีและเริ่มต้นทำงานใหม่ทุก ๆ ครั้งที่มีการขอบริการ ซึ่งทำให้การบริการของเว็บเซิร์ฟเวอร์จะไม่มีประสิทธิภาพเมื่อต้องมีการขอบริการจำนวนมาก ๆ

standalone คือการให้บริการโดยให้โปรแกรมอาปาเช่เว็บเซิร์ฟเวอร์จัดการควบคุมและบริหารโปรเซสที่ใช้ในการให้บริการเอง เป็นรูปแบบที่มีประสิทธิภาพสูงและมีความยืดหยุ่นสูงในการที่ปรับเปลี่ยนคุณสมบัติต่าง ๆ เพื่อให้มีประสิทธิภาพสูงขึ้น

- **Port {PortNumber}**

Port เป็นการกำหนดหมายเลขพอร์ตที่จะใช้ในการบริการ โดยปกติคือหมายเลข 80

- **HostnameLookups {on,off,double}**

HostnameLookups เป็นการกำหนดให้เว็บเซิร์ฟเวอร์ทำการตรวจหมายเลขไอพีของเครื่องคอมพิวเตอร์ที่ขอบริการหรือไม่ ค่าเดิมจะเป็น off โดยจะมีตัวเลือกด้วยกันดังต่อไปนี้

– off คือการกำหนดให้เว็บเซิร์ฟเวอร์ไม่ต้องทำการตรวจหมายเลขไอพีของเครื่องคอมพิวเตอร์ที่ขอบริการ

– on คือการให้เว็บเซิร์ฟเวอร์ต้องทำการตรวจหมายเลขไอพีของเครื่องคอมพิวเตอร์ที่ขอบริการกับโดเมนเนมเซิร์ฟเวอร์ (DNS) ว่าหมายเลขไอพีนั้นมีการบันทึกชื่อไว้หรือไม่ ถ้าไม่จะไม่มีการให้บริการถ้ามีการกำหนดขอบเขตการให้บริการไว้

– double คือการให้เว็บเซิร์ฟเวอร์ต้องทำการตรวจหมายเลขไอพีของเครื่องคอมพิวเตอร์ที่ขอบริการกับโดเมนเนมเซิร์ฟเวอร์ (DNS) ว่าหมายเลขไอพีนั้นมีการบันทึกชื่อไว้หรือไม่ หลังจากนั้นก็จะนำชื่อที่ตรวจสอบได้ทำการตรวจสอบไปที่โดเมนเนมเซิร์ฟเวอร์อีกครั้งว่าได้หมายเลขไอพีที่ตอบกลับตรงกับหมายเลขไอพีแรกหรือไม่ ถ้าไม่จะไม่มีการให้บริการถ้ามีการกำหนดขอบเขตการให้บริการไว้

- **User {UserID, UserName}**

User เป็นการกำหนดชื่อผู้ใช้ให้เว็บเซิร์ฟเวอร์โดยชื่อผู้ใช้ที่กำหนดจะต้องมีอยู่ในไฟล์ /etc/passwd

- **Group {GroupID, GroupName}**

Group เป็นการกำหนดชื่อกลุ่มผู้ใช้ให้เว็บเซิร์ฟเวอร์โดยชื่อกลุ่มผู้ใช้ที่กำหนดจะต้องมีอยู่ในไฟล์ /etc/group

- **ServerAdmin {EMailAddress}**

ServerAdmin เป็นการกำหนดอีเมล์ของผู้ดูแลเว็บเซิร์ฟเวอร์ เพื่อที่ว่าเวลาเกิดมีปัญหาต่าง ๆ กับเว็บเซิร์ฟเวอร์ เว็บเซิร์ฟเวอร์จะทำการรายงานปัญหาต่าง ๆ ที่เกิดขึ้นผ่านทางอีเมล์แทน

- **ServerRoot {DirectoryName}**

ServerRoot เป็นการกำหนดไดเรคทอรีที่เว็บเซิร์ฟเวอร์จะใช้ในการทำงานปกติจะเป็น /etc/httpd โดยไดเรคทอรีที่กำหนดจะประกอบด้วยไดเรคทอรีย่อย conf และ logs

- **DocumentRoot {DirectoryName}**

DocumentRoot เป็นการกำหนดตำแหน่งที่ของเอกสารที่ต้องการให้เว็บเซิร์ฟเวอร์ใช้ในการให้บริการ เช่น DocumentRoot /usr/htdocs เมื่อผู้ใช้ระบุในเว็บเบราเซอร์ว่า http://www.ku.ac.th/index.html เว็บเซิร์ฟเวอร์จะรับข้อความและตีความว่าให้อ่านเอกสารที่ผู้ใช้ต้องการจากไฟล์นี้แทน /usr/htdocs/index.html

- **ErrorDocument {ErrorCode} {FileName}**

ErrorDocument เป็นการกำหนดเอกสารที่ต้องการให้แสดงแทนเมื่อเกิดปัญหาที่เว็บเซิร์ฟเวอร์ เราสามารถกำหนดให้แสดงเอกสารที่แตกต่างกันเมื่อเกิดปัญหาแตกต่างกันได้ โดยให้ดูตามหมายเลขของ ErrorCode ที่เกิดขึ้น

- **IdentityCheck {on,off}**

IdentityCheck เป็นการกำหนดเอกสารที่ต้องการให้แสดงแทนเมื่อเกิดปัญหาที่เว็บเซิร์ฟเวอร์ เราสามารถกำหนดให้แสดงเอกสารที่แตกต่างกันเมื่อเกิดปัญหาแตกต่างกันได้ โดยให้ดูตามหมายเลขของ ErrorCode ที่เกิดขึ้น

การกำหนดคุณสมบัติไฟล์ access.conf

- **order {allow, deny}, {deny, allow}**

Order ใช้กำหนดลำดับการตรวจสอบการยินยอมให้เข้าถึงเอกสารจากเครื่องคอมพิวเตอร์ใด ๆ เช่น

order allow, deny ระบบจะทำการตรวจสอบคุณสมบัติ allow ก่อนคุณสมบัติ deny

order deny, allow ระบบจะทำการตรวจสอบคุณสมบัติ deny ก่อนคุณสมบัติ allow

- **deny from host ...**

deny เป็นการกำหนดให้เว็บเซิร์ฟเวอร์ป้องกันการเข้าถึงเอกสารต่าง ๆ จากเครื่องคอมพิวเตอร์ที่ระบุไว้ เช่น

deny from ku.ac.th

เป็นการป้องกันไม่ให้เครื่องคอมพิวเตอร์ทุก ๆ เครื่องที่อยู่ภายใต้โดเมน ku.ac.th. เข้าถึงเอกสารต่าง ๆ

deny from 158.108.2.71

เป็นการป้องกันไม่ให้เครื่องคอมพิวเตอร์ที่มีหมายเลขไอพี 158.108.2.71 เข้าถึงเอกสารต่าง ๆ

allow from host ...

allow เป็นการกำหนดให้เว็บเซิร์ฟเวอร์ยินยอมเข้าถึงเอกสารต่าง ๆ จากเครื่องคอมพิวเตอร์ที่ระบุไว้ เช่น

allow from ku.ac.th

เป็นการยินยอมให้เครื่องคอมพิวเตอร์ทุก ๆ เครื่องที่อยู่ภายใต้โดเมน ku.ac.th. เข้าถึงเอกสารต่าง ๆ

allow from 158.108.2.71

เป็นการยินยอมให้เครื่องคอมพิวเตอร์ที่มีหมายเลขไอพี 158.108.2.71 เข้าถึงเอกสารต่าง ๆ

สารพันปัญหาถามตอบกับ Linux

ถาม อาปาเช่เว็บเซิร์ฟเวอร์มีปัญหาปี 2000 หรือไม่

ตอบ มีสภาวะ 3 อย่างที่ใช้พิจารณาปัญหาปี ค.ศ. 2000 กับอาปาเช่คือ

1. ส่วนของโค้ดในโปรแกรมอาปาเช่นั้นจัดการเกี่ยวกับวันและเวลาในมาตรฐานของระบบเวลาในยูนิกซ์ซึ่งตัวเลขของปีเป็นตัวเลข 4 หลัก

2. รูปแบบของวันเวลาที่โปรโตคอล http และโปรโตคอลอื่น ๆ ที่อาปาเช่ใช้ รูปแบบของเวลามีอยู่ 3 แบบคือ

2.1 มาตรฐาน RFC822 ซึ่งปรับปรุงเป็น RFC 1123 มีรูปแบบเป็น

Sun,13 Nov 1994 08:59:59 GMT

2.2 มาตรฐาน RFC 850 และ RFC 1036 มีรูปแบบเป็น

Sunday,13-Nov-94 08:59:59 GMT

2.3 รูปแบบของ asctime() ในมาตรฐาน ANSI

Sun Nov 13 08:59:59 1994 GMT

ปัญหาปี ค.ศ. 2000 ที่เกิดนั้นเนื่องจากโปรแกรมเบราเซอร์รุ่นเก่าใช้มาตรฐาน RFC 850 ซึ่งกำหนดตัวเลขปี ค.ศ. เป็นเลข 2 หลัก แต่คาดว่าเว็บเบราเซอร์นั้นจะถูกแทนที่ด้วยรุ่นใหม่ที่มีรูปแบบของปีเป็นเลข 4 หลักแทน

3. ขึ้นอยู่กับระบบปฏิบัติการที่ใช้งานว่ามีปัญหากับปี 2000 หรือไม่ Linux และยูนิกซ์นั้นมีรูปแบบของการจัดเก็บเวลาภายในเป็นแบบ 32 บิตซึ่งไม่มีปัญหากับปี 2000 แต่จะเกิดปัญหาเมื่อปี 2038 แทน ซึ่ง ณ เวลานั้นโปรแกรมก็เลิกใช้หรือหมดอายุแล้ว (software life cycle)

ถาม อาปาเช่รุ่นใหม่ล่าสุดนั้นเป็นรุ่นที่เท่าไร

ตอบ อาปาเช่เป็นรุ่นที่ 1.3.6 ซึ่งเป็นรุ่นที่พัฒนาและแก้ไขข้อบกพร่องจากอาปาเช่รุ่นที่ 1.3.4 (รุ่นที่ 1.3.5 เป็นรุ่นที่ไม่มีการเผยแพร่) ส่วนการที่ผู้ใช้จะเปลี่ยนอาปาเช่จากรุ่น 1.3.4 เป็น 1.3.6 หรือไม่นั้นขึ้นอยู่กับระบบปฏิบัติการที่ใช้ เช่น หากระบบ Unix หรือ Linux นั้นควรเปลี่ยน แต่ถ้าหากใช้งานบนระบบ Windows คงต้องพิจารณา เนื่องจากแฟ้ม htpasswd ในรุ่นที่ 1.3.4 ไม่สามารถใช้ได้ในรุ่นที่ 1.3.6 หากไม่มีการอัพเดต

สรุป

อาปาเช่เป็นเว็บเซิร์ฟเวอร์ที่นิยมใช้กันมากบนระบบปฏิบัติการหลายประเภท เนื่องจากความสามารถและคุณสมบัติที่มีการปรับเปลี่ยนให้ทันสมัยอย่างต่อเนื่องของอาปาเช่ สำหรับเรื่องราวของ Linux ในฉบับนี้คงต้องลากันแล้ว ค่อยพบกันใหม่ในฉบับหน้านะครับ